Automation and Remote Control, Vol. 55, No. 8, 1994

STOCHASTIC SYSTEMS

SECOND-ORDER CONDITIONS OF IMPROVEMENT IN NONLOCAL OPTIMIZATION METHODS BASED ON POTENTIAL THEORY

A. I. Kaplinskii and A. I. Propoi

UDC 517.977.5

The problem of minimizing a function of several variables is considered. This problem is replaced by an equivalent randomized problem in infinite-dimensional space. Constructive characteristics of the variation of the functional of the randomized problem are obtained. On this basis, second-order optimization methods (in particular, the Newton method) that require only values of the objective function for their implementation are developed.

1. INTRODUCTION

It is well known that the second-order methods, which use the quadratic approximation of the objective function, are one of the most effective optimization methods [1]. However, these methods are generally developed for "good" functions (convex, sufficiently smooth, etc.). In the present work we develop second-order methods for minimizing a function of several variables about which only its values are known. These methods are based on replacing the original problem by a randomized problem that, in turn, makes it possible to determine a potential function which is a solution of a boundary-value problem in mathematical physics and serves as a model of extremal properties of the original function. This approach was considered in [2]. In the present work, which is a continuation of [2], we obtain constructive characteristics of the variation of the objective functional of the randomized problem and on this basis develop second-order methods (in particular, the Newton method) that require only values of the objective function for their implementation.

2. STATEMENT OF THE PROBLEM

We consider the problem of unconstrained minimization

$$f(x) \to \min_{x \in R^n}, \tag{1}$$

where the function $f : R^n \to R$. We assume that generally only values of the function f can be measured at different points x from R^n. This problem is replaced by a randomized problem

$$F(X) = E[f(X)] \to \min_{X \in \{X\}}, \tag{2}$$

where X is a random vector with values in R^n and with the density function $p(x)$ understood, possibly, in the generalized sense.

Problem (2) is equivalent to problem (1) in the sense that any realization of the random vector X^*, where X^* is a solution of problem (2), that has a nonzero probability will be a solution of problem (1) [2].

Voronezh State University. Institute for System Analysis, Russian Academy of Sciences, Moscow. Translated from Avtomatika i Telemekhanika, No. 8, pp. 104–117, August, 1994. Original article submitted November 22, 1993.

 0005-1179/94/5508-1170$12.50

Let $X_\varepsilon = X + \varepsilon Y$, where the random vector Y with the joint probability density function $p(x, y)$ determines the direction of the variation of the state X, $\varepsilon \geqslant 0$. Correspondingly,

$$F(X + \varepsilon Y) = F(X) + \varepsilon \frac{d}{d\varepsilon}[F(X + \varepsilon Y)]_{\varepsilon=0} + \frac{1}{2}\varepsilon^2 \left[\frac{d^2}{d\varepsilon^2} F(X + \varepsilon Y)\right]_{\varepsilon=0} + o(\varepsilon^2). \tag{3}$$

The direction Y satisfying the inequality

$$F(X + \varepsilon Y) - F(X) \leqslant 0 \tag{4}$$

for $0 < \varepsilon \leqslant \varepsilon_1$ will be called the improving direction.

The first-order variation in (3) and the constructive conditions of improvement of the current state following from it are given in [2, 3]. The goal of this work is to obtain the second-order variation and to study, on this basis, the structural components of nonlocal search.

3. VARIATION OF THE FUNCTIONAL AND CONDITIONS OF IMPROVEMENT

By using the conditional probability function

$$p(x, y) = p(x)p(y|x) \tag{5}$$

we introduce the following characteristics of the random vector Y:

$$\bar{y}(x) = E[Y|X = x] = \int_{R^n} y p(y|x)\, dy, \tag{6}$$

$$K(x) = E\left[(Y - \bar{y}(x))(Y - \bar{y}(x))^T | X = x\right] = \int_{R^n} (y - \bar{y}(x))(y - \bar{y}(x))^T p(y|x)\, dy, \tag{7}$$

where E is the symbol of mathematical expectation and T denotes transposition.

Proposition 1. The variation (3) of the functional $F(X)$ defined in (2) is represented in the form

$$F(X + \varepsilon Y) - F(X) = -\varepsilon \int_{R^n} (f(x) - c)\mathrm{div}[p(x)\bar{y}(x)]\, dx + \tag{8}$$

$$\frac{1}{2}\varepsilon^2 \int_{R^n} (f(x) - c)\mathrm{div}\left\{\bar{y}(x)\mathrm{div}[p(x)\bar{y}(x)] + \nabla^T \bar{y}(x) p(x)\bar{y}(x)\right\} dx +$$

$$\frac{1}{2}\varepsilon^2 \int_{R^n} (f(x) - c)\mathrm{div}h(x)\, dx + o(\varepsilon^2),$$

where $\nabla\bar{y}(x)$ is a matrix composed of the columns $\nabla\bar{y}_i(x)$ $(i = 1, \ldots, n)$; $h(x)$ is a vector with the coordinates $\mathrm{div}[p(x)K_i(x)]$, $K_i(x)$ $(i = 1, \ldots, n)$ is the i-th column of the covariance matrix $K(x)$.

The proof of Proposition 1 is given in the Appendix.

Thus, in expression (8) for the variation of functional (3) only the characteristics (6) and (7) of the random vector Y are present. Since these characteristics can be chosen independently, it follows straightforwardly from (4) and (8)

Proposition 2. A random vector Y is an improving direction if $\bar{y}(x)$ and $K(x)$ defined in (6) and (7) satisfy the inequalities

$$\int_{R^n} (f(x) - c)\mathrm{div}[p(x)\bar{y}(x)]\, dx - \tag{9}$$

$$\frac{1}{2}\varepsilon \int_{R^n} (f(x) - c)\mathrm{div}\{\bar{y}(x)\mathrm{div}[p(x)\bar{y}(x)] + \nabla^T \bar{y}(x) p(x)\bar{y}(x)\}\, dx \geqslant 0,$$

$$\int_{R^n} (f(x) - c)\mathrm{div}h(x)\, dx \leqslant 0. \tag{10}$$

We shall assume that the inequalities (9) and (10) are fulfilled independently, on different stages of nonlocal search. In this case it is sufficient to set

$$\operatorname{div}\{p(x)\bar{y}(x) - \frac{1}{2}\varepsilon\bar{y}(x)\operatorname{div}[p(x)\bar{y}(x)] - \frac{1}{2}\varepsilon p(x)\nabla^T\bar{y}(x)\bar{y}(x)\} = (f(x) - c_1)p_U(x), \tag{11}$$

$$\operatorname{div}h(x) = -(f(x) - c_2)p_V(x), \tag{12}$$

where the functions $p_U(x)$, $p_V(x)$ are nonnegative and have, possibly, the meaning of the density functions of random vectors U and V, respectively. These functions can be chosen. Therefore, $\bar{y}(x)$ and $K(x)$ are defined up to controls — the random vectors U and V.

We have, from differentiation with respect to ε the identity

$$\int_{R^n} p(x - \varepsilon y, y)\,dy = 1,$$

and from the relations (11) and (12), that

$$c_1 = \int_{R^n} f(x)p_U(x)\,dx = E[f(U)], \tag{13}$$

$$c_2 = \int_{R^n} f(x)p_V(x)\,dx = E[f(V)]. \tag{14}$$

Substituting (11) and (12) in (8) we obtain

$$\delta_U F(X) = \int_{R^n} (f(x) - c_1)^2 p_U(x)\,dx = E[f(U) - c_1]^2, \tag{15}$$

$$\delta_V F(X) = \int_{R^n} (f(x) - c_2)^2 p_V(x)\,dx = E[(f(V) - c_2]^2, \tag{16}$$

where by $\delta_U F(X)$ and $\delta_V F(X)$ we denote increments of the functional $F(X)$ given by the distributions $p_U(x)$ and $p_V(x)$, respectively.

Thus, the improving direction Y is defined by the vector field $\bar{y}(x)$ and the matrix $K(x)$ which depend on the choice of the random vectors U and V and must satisfy Eqs. (11) and (12); the corresponding (improving) increments of the functional $F(X)$ have the meaning of the variance of the objective function f over the distributions of U and V.

The relations (11) and (12), however, are not constructive because they do not define uniquely the characteristics of Y introduced above. To specify them we make use of the partitition of the vector field $\bar{y}(x)$

$$p(x)\bar{y}(x) = \nabla\varphi(x) + w(x), \qquad \operatorname{div}w(x) = 0. \tag{17}$$

(We partitition the vector field $h(x)$ in a similar way.)

These partitions can always be made for finite or vanishing-at-infinity vector fields; for the additional condition: $\varphi(x) \to 0$, $\|x\| \to \infty$ they are unique [4].

We consider at first the relation (11). It contains a small parameter ε, so that $\bar{y}(x) = g(x, \varepsilon)$. We restrict ourselves to the case in which $g(x, \varepsilon)$ admits the expansion in powers of ε:

$$g(x, \varepsilon) = g_0(x) + \varepsilon g_1(x) + \ldots. \tag{18}$$

Substituting (18) in (11) and equating members with the same power of ε, we obtain

$$\operatorname{div}[p(x)g_0(x)] = (f(x) - c)p_U(x), \tag{19}$$

$$\operatorname{div}\left\{p(x)g_1(x) - \frac{1}{2}\sum_{i=1}^{n}\frac{\partial[p(x)g_{0i}(x)g_0(x)]}{\partial x_i}\right\} = 0. \tag{20}$$

We now make use of the decomposition (17) of the vector field $\bar{y}(x)$, assuming, for simplicity, that $w(x) \equiv 0$. Then as a solution of (19) we can take the Newtonian field

$$p(x)g_0(x) = \nabla\varphi(x), \tag{21}$$

where the potential $\varphi(x)$ satisfies the Poisson equation in R^n,

$$\Delta\varphi(x) = (f(x) - c)p_U(x), \tag{22}$$

where for $n > 2$ [5] we have

$$\varphi(x) = \int_{R^n} G(x,\xi)(f(\xi) - c)p_U(\xi)\,d\xi = -\frac{1}{\omega^n(n-2)}\int_{R^n}\frac{f(\xi)-c}{\|x-\xi\|^{n-2}}p_U(\xi)\,d\xi, \tag{23}$$

$$\nabla\varphi(x) = \frac{1}{\omega^n}\int_{R^n}\frac{x-\xi}{\|x-\xi\|^n}(f(\xi) - c)p_U(\xi)\,d\xi, \tag{24}$$

and the value $c = c_1$ is defined by the equality (13).

As a particular solution of (20) we take a function $g_1(x)$ that satisfies the relation

$$p(x)g_1(x) = \frac{1}{2}\sum_{i=1}^{n}\frac{\partial[p(x)g_{0i}(x)g_0(x)]}{\partial x_i} = \frac{1}{2}g_0(x)\mathrm{div}[p(x)g_o(x)] + \frac{1}{2}\sum_{i=1}^{n}p(x)g_{0i}(x)\frac{\partial g_0(x)}{\partial x_i}. \tag{25}$$

It follows from (21) that

$$g_{0i}(x) = \frac{1}{p(x)}\frac{\partial\varphi(x)}{\partial x_i},$$

$$\frac{\partial g_{0i}(x)}{\partial x_i} = \frac{1}{p(x)}\frac{\partial^2\varphi(x)}{\partial x_i\partial x_j} - \frac{1}{p^2(x)}\frac{\partial\varphi(x)}{\partial x_i}\frac{\partial p(x)}{\partial x_j}.$$

Substituting these expressions in (25) we obtain, in matrix form, that

$$g_1(x) = \frac{1}{2}\left[(f(x) - c)I + \frac{1}{p(x)}\nabla^2\varphi(x) - \frac{1}{p^2(x)}\nabla\varphi(x)\nabla^T p(x)\right]g_0(x), \tag{26}$$

where I is the identity matrix.

Thus, restricting the expansion by the terms obtained we can represent the vector field $\bar{y}(x)$ in the form

$$\bar{y}(x) = \left\{I + \frac{\varepsilon}{2}\left[(f(x) - c)I + \frac{\nabla^2\varphi(x)}{p(x)} - \frac{\nabla\varphi(x)\nabla^T p(x)}{p^2(x)}\right]\right\}\frac{\nabla\varphi(x)}{p(x)}. \tag{27}$$

The relations obtained allow us to formulate

Proposition 3. Let, in the random process

$$X_{N+1} = X_N + \alpha_N\bar{y}_N(X_N), \qquad N = 0, 1, \ldots, \tag{28}$$

the vector field $\bar{y}_N(x)$ satisfy (27) for $p(x) = p_N(x)$, where $p_N(x)$ is the density function of the random vector X_N, and let the function $\varphi(x) = \varphi_N(x)$ be the solution (23) of the Poisson equation (22), where $c = c_N$ is determined from (13) if $p_U(x) = p_{U_N}(x)$; then $F(X_{N+1}) \leqslant F(X_N)$ for an appropriate choice of the length step $\alpha_N > 0$.

If, in (27), we set $\varepsilon = 0$, then $\bar{y}(x) = \nabla\varphi(x)/p(x)$ (see (21)) and we obtain the conditions of improvement of the first order leading, in particular, to the gradient descent method (with respect to the potential function φ, but not to the original function f). These methods and the corresponding structural components of nonlocal search are considered in [2, 3, 6, 7]. In these papers it is shown that the potential function φ serves as a substitute, a model of extremal properties of the objective function f. Therefore, in addition to the gradient descent method we can consider the Newton method for the potential function φ

$$X_{N+1} = X_N - \left[\nabla^2\varphi(X_N)\right]^{-1}\nabla\varphi(X_N) \tag{29}$$

and use it, as well as (28) and (27), for a basis of developing the second-order nonlocal search algorithms. In both cases (e.g., both for (28) and (29)) it is necessary to evaluate the Hessian of the potential function $\nabla^2\varphi$ for the implementation of these methods.

Proposition 4. The matrix of second derivatives of the potential function admits the representation

$$\nabla^2\varphi(x) = \frac{\tilde{f}(x)p_U(x)}{n}I + E_U[\nu(U,x)(I - \theta(U,X)\theta^T(U,x))], \tag{30}$$

where E_U is the symbol of mathematical expectation over the random vector U, $\theta(U,x)$ is a random normalized direction from the point x:

$$\theta(U,x)=\frac{U-x}{\|U-x\|}, \tag{31}$$

$$\tilde{f}(x)=f(x)-c, \tag{32}$$

$$\nu(U,x)=\chi_{S_\rho(x)}\frac{f(U)-f(x)\dfrac{p_U(x)}{p_U(U)}}{\omega_n\|U-x\|^n}+\chi_{R^n\setminus S_\rho(x)}\frac{\tilde{f}(U)}{\omega_n\|U-x\|^n}. \tag{33}$$

where ω_n is the area of the unit sphere in R^n, $S_\rho(x)$ is the ball in R^n with the center x and the radius ρ: $S_\rho(x)=\{u|\ \|u-x\|^2\leqslant\rho^2\}$, χ_D is the characteristic (indicator) function of the set D.

The proof of Proposition 4 is given in the Appendix.

Making use of the expression (30) for the matrix $\nabla^2\varphi(x)$ we represent $\bar{y}(x)$ in the form

$$\bar{y}(x)=\frac{1}{p(x)}R(x)\nabla\varphi(x), \tag{34}$$

where the operator $R(x)$ is given by the matrix

$$R(x)=\left[1+\frac{\varepsilon}{2}\left(\tilde{f}(x)\left(1+\frac{1}{n}\right)+\frac{1}{p(x)}E[\nu(U,x)]-\frac{\nabla^T p(x)\nabla\varphi(x)}{p^2(x)}\right)\right]I- \tag{35}$$

$$-\frac{\varepsilon n}{2p(x)}E[\nu(U,x)\theta(U,x)\theta^T(U,x)].$$

Denote

$$\gamma(x)=1+\frac{\varepsilon}{2}\left[\tilde{f}(x)\left(1+\frac{1}{n}\right)+\frac{E[\nu(U,x)]}{p(x)}-\frac{\nabla^T p(x)\nabla\varphi(x)}{p^2(x)}\right]. \tag{36}$$

We can always assume that $\gamma>0$ for sufficiently small ε. Then (35) is rewritten as

$$R=\gamma E\left[I-\frac{\varepsilon n}{2\gamma p}\nu\theta\theta^T\right]. \tag{37}$$

Thus, the direction of motion Y, given by the vector field $\bar{y}(x)$ (27), is represented in the form (34), (37) and (36).

The discussion of the relations obtained is given in the next section. Here we make only general remarks.

Remark 1. If we modify in an appropriate way the right-hand side of the Poisson equation (22), then we can make the matrix $\nabla^2\varphi$ positive definite in the vicinity of the (global) minimum f [8]. In this case we can evaluate the rate of convergence to the global minimum for the algorithm (29) [8].

Remark 2. The use of characteristics of the expansion of more than second order gives a new quality to the problem (requiring a generalization of the notion of integral) because even for the representation of $\nabla^2\varphi$ the notion of singular integral was used (see the Appendix).

Remark 3. Inequality (10), which determines the choice of the covariance matrix $K(x)$, can be handled in a similar way. Indeed, representing the vector field $h(x)$ as the sum of potential and "nondivergent" fields and substituting this decomposition into (12), we obtain up to the nondivergent component that (cf. (23) and (24))

$$h(x)=-\int_{R^n}\nabla_x G(x,\xi)(f(\xi)-c_2)p_V(\xi)\,d\xi,$$

where the constant c_2 is defined from (14). This reasoning can be repeated because $h_i(x)=\operatorname{div}[p(x)K_i(x)]$; as a result

$$p(x)K(x)=-\int_{R^n}\int_{R^n}\nabla_x G(x,\eta)\nabla_\eta^T G(\eta,\xi)(f(\xi)-c)p_V(\xi)\,d\xi\,d\eta. \tag{38}$$

The algorithm for updating the covariance matrix $K(x)$ can be developed if we use, for example, the normal distributions to specify random vectors X_N [6, 7].

Remark 4. The principal feature that characterizes the structure of the vector field $\bar{y}(x)$ (as well as the structure of the Hessian of the potential function) is the presence of a unified operator of the statistical processing of information. The matrix of this operator can be written in the form

$$R(x) = I + n\alpha(x)E[\nu(U,x)\theta(U,x)\theta^T(U,x)], \tag{39}$$

where information on the properties of the objective function is accumulated in the structure of the function $\nu(U,x)$ (33).

The transformation given by (39) combines the operations of dilatation and reflection. We recall that the operator of space dilatation is given by $R_\beta(\theta) = I + (\beta - 1)\theta\theta^T$ [9]. It multiplies on $\beta \geqslant 0$ any vector that is collinear to θ and does not change vectors that are orthogonal to θ. If $\beta < 0$ it reflects vectors over the hyperplane that is orthogonal to θ (with the corresponding change of the length). We emphasize the dependence of the operator (39) on the dimension n of the search space and its relation to the structure of covariance matrices.

4. THE STRUCTURE OF NUMERICAL METHODS

Above obtained formulas serve as the basis for the development of second-order numerical methods of nonlocal search. As was mentioned, these methods can be implemented in two main ways: the specification of the dynamics of random vectors (28) by the vector field $\bar{y}_N(x)$, which is one of two characteristics of the random direction Y_N of changing X_N; the implementation of the Newton method (29) for the potential function.

Before discussing the second-order methods, we consider the major specific features of implementing the first-order methods. As we noted, these methods follow from (27) and (28) if $\varepsilon = 0$ and lead to the gradient descent with respect to the potential function. The expression for the gradient of the potential function in the integral form is given in (24). Using the approach accepted in adaptive (probabilistic iterative) methods [10, 11], we assume that the estimate of $\nabla\varphi$ is carried out over realizations u_N of the random vector U developed in time. As a result, we obtain the gradient descent method (with respect to the potential function) written in the following way:

$$x_{N+1} = x_N - \alpha_N \frac{f(u_N) - c_N)}{\omega_n \|u_N - x_N\|^{n-1}} \theta(u_N, x_N), \tag{40}$$

where $\theta(u_N, x_N) = (u_N - x_N)/\|u_N - x_N\|$ (see (31)).

The convergence of the probabilistic iterative methods has been well studied [10, 11]. A specific feature of the case considered is the presence of the pole in the kernel of transformation (24) which insures instability, pushing the search point out of the nonperspective domain (in particular, out of the zones of local extrema of f).

In the perspective domain $\{x | f(x) \leqslant c\}$ the localization of the optimal solution is performed. In this case, according to (40), the shift of x_N in the direction $\theta(u_N, x_N)$ will be larger, the stronger the deviation of $f(u_N)$ from the level c_N. For the evaluation of the value c_N defined by (13) we can also use the realizations of the random vector U in an iterative procedure similar to (40).

Thus, even in the first order methods, which realize gradient descent with respect to the potential function and use only the values of the objective function f, the nonlocal character of the search is insured (for details, see [2, 6]).

However, the process of nonlocal search may be complicated by the ravine behavior of the objective function, which can be inherited by the potential function. From this the necessity of second-order methods, the consideration of which we are passing to, is obvious.

We assume that we have at our disposal K realizations of the random vector U for the evaluation of values entering the representation (34), (37), and (36) of the vector field $\bar{y}(x)$. Then

$$\bar{y}(x) = \frac{\gamma}{p}\left[I - \frac{\varepsilon n}{2K\gamma p}\sum_{i=1}^{K} \nu_i(x)\theta_i(x)\theta_i^T(x)\right]\nabla\varphi(x), \tag{41}$$

where $\nu_i(x) = \nu(u_i, x), \theta_i(x) = \theta(u_i, x)$. (We use the same notation for the vector field (34) and its estimate (41).)

The latter expression can be written in the form

$$\bar{y}(x) = \frac{\gamma}{p}\left[\nabla\varphi - \frac{\varepsilon n}{2K\gamma p}\sum_{i=1}^{K} \nu_i(\theta_i, \nabla\varphi)\theta_i\right]. \tag{42}$$

We can see from the repesentation (42) that it is composed of two terms: the vector $\nabla\varphi$ that defines the direction of maximizing φ is supplemented by realizations of the random vector $\theta_i (i, \ldots, K)$ with certain weight coefficients. We explain the meaning of these coefficients.

The random value $\nu(U, x)$ defined in (33) consists of two complementary terms. The first term occurs in ν_i when the realization u_i gets into the ball S_ρ. In this term the deviations of the function f from the level c at the points u_i and x are compared. In this case $\nu_i = \nu(u_i, x) < 0$ when $f(u_i) < f(x)$ and $\nu_i \geqslant 0$ otherwise. In the second term, which can be made dominant by the choice of ρ, the sign of ν_i is defined by the deviation of the value $f(u_i)$ from the level c. In this case $\nu_i < 0$ when the realization u_i is perspective.

The inner product $(\theta_i, \nabla\varphi)$ defines the perspectiveness of the realization $\theta_i = \theta(u_i, x)$ with respect to $\nabla\varphi$: $(\theta_i, \nabla\varphi) > 0$ if the vector θ_i has an acute angle with $\nabla\varphi$. Taking into consideration that the magnitude $\gamma(x)$ in (42) can be considered positive (see (36)), we obtain that the vector field $\bar{y}(x)$ is represented in the form $\bar{y}(x) \sim \nabla\varphi + \sum_i \alpha_i\theta_i$, where $\alpha_i > 0$ for perspective realizations θ_i.

To form the vector field $\bar{y}(x)$ it is convenient to use an iterative (adaptive) procedure setting

$$\bar{y}(x) = \frac{\gamma(x)}{p(x)}\left[\nabla\varphi(x) + \sigma_i(x) - 1)\theta_i(x)\right], \tag{43}$$

$$\sigma_{i+1}(x) = \begin{cases} a_i(x), & \text{if} \quad a_i(x) > 1, \\ 1, & \text{if} \quad a_i(x) \leqslant 1, \end{cases} \tag{44}$$

where

$$a_i(x) = \sigma_i(x) - \frac{\varepsilon n}{2K\gamma p}\nu_i(\theta_i, \nabla\varphi).$$

The magnitude σ_i ensures the accumulation ("memory") of perspective directions θ_i, realizing the updating of the coefficient before θ_i over a definite number of iterations.

We obtain from (43) and (44) that

$$(\sigma_i - 1)\theta_i = -\frac{\varepsilon n\nu_i}{2K\gamma p}(\theta_i, \nabla\varphi)\theta_i = (\beta_i - 1)(\theta_i, \nabla\varphi)\theta_i,$$

where

$$\beta_i = 1 - \frac{\varepsilon n\nu_i}{2K\gamma p}.$$

If the vector θ_i has an acute angle with $\nabla\varphi$: $(\theta_i, \nabla\varphi) > 0$, then we obtain $\beta_i \geqslant 1$, i.e., $\sigma_i \geqslant 1$, taking into account that $\nu_i \leqslant 0$ for the perspective realization θ_i.

Thus, for perspective realizations θ_i ($(\theta_i, \nabla\varphi) > 0$, $\theta_i \leqslant 0$) the magnitude σ_i increases, amplifying the "trust" to the direction θ_i. If $(\theta_i, \nabla\varphi) < 0$ and $\alpha_i < 1$ (an uncertain situation in the process of search), then it makes sense to set $\sigma_{i+1} = 1$, switch off the second term in (44) and together with it the memory about the previous steps.

We now analyze the role of the operator of statistical processing of information introduced in (39). First, we consider the structure of the Hessian of the potential function (30). For this we rewrite (30) in the form (cf. (39))

$$\nabla^2\varphi(x) = \tilde{\gamma}(x)\{I - (n/\tilde{\gamma}(x))E[\nu(U, x)\theta(U, x)\theta^T(U, x)]\}, \tag{45}$$

where

$$\tilde{\gamma}(x) = \frac{\tilde{f}(x)p(x)}{n} + E[\nu(U, x)]. \tag{46}$$

We shall use K realizations for the statistical estimate of the Hessian (45). Then, if, in addition, the vectors $\theta_i = \theta(u_i, x)$, $i = 1, \ldots, K \leqslant n$, are orthogonal, the matrix $\nabla^2\varphi$ can be represented as the product of K elementary operators $\tilde{R}_i$:

$$\nabla^2\varphi = \tilde{\gamma}\left[I - \frac{n}{\tilde{\gamma}K}\sum_{i=1}^{K}\nu_i\theta_i\theta_i^T\right] = \tilde{\gamma}\prod_{i=1}^{K}\tilde{R}_i,$$

where

$$\tilde{R}_i = I + (\tilde{\beta}_i - 1)\theta_i\theta_i^T, \tag{47}$$

$$\tilde{\beta}_i = 1 - n\nu_i/\gamma K \tag{48}$$

and the same notation is used both for the matrix $\nabla^2\varphi$ and its estimate.

The operator $\tilde{R}_i$ is easily inverted [9]

$$\tilde{R}_i^{-1} = I + ((1/\tilde{\beta}_i - 1)\theta_i\theta_i^T.$$

Correspondingly, the matrix $[\nabla^2\varphi]^{-1}$ takes the form

$$[\nabla^2\varphi]^{-1} = (1/\widetilde{\gamma})\prod_{i=1}^{K}[I + ((1/\widetilde{\beta}_i) - 1)\theta_i\theta_i^T].$$

We limit ourselves by a single realization ($K = 1$) of the random vector θ_i (assuming, as before, that the process of accumulation of realizations is developed in time) and analyze the effect of this matrix on the vector $\nabla\varphi$.

We assume that the sign of $\widetilde{\gamma}$ (46) is determined by the realization $\nu_i = \nu(u_i, x)$. As was mentioned before, the magnitude ν_i characterizes the perspectiveness of the realization u_i, both with respect to the level c (the second term in (33)) and to the value of f at the search point x (the first term in (33)).

Let the realization u_i be perspective, i.e., $\nu_i < 0$. Then $1/\beta_i < 0$ for $n > 2$, i.e., the operator $\widetilde{R}_i^{-1}$ reflects the vector $\nabla\varphi$ over the hyperplane orthogonal to θ_i. If, in addition, $\widetilde{\gamma} < 0$, then the half-space that contains $\nabla\varphi$ is considered perspective, and the vector reflected is taken with the opposite sign, which returns it to the perspective half-space. If $\widetilde{\gamma} > 0$, then the half-space that contains $\nabla\varphi$ is nonperspective and the vector $\nabla\varphi$ must be moved to the opposite half-space.

We also note that the ratio $\nu_i/\widetilde{\gamma}K$ in the coefficient $\widetilde{\beta}$ is close to unity and, consequently, the operator $\widetilde{R}_i$ is close to the conventional operator of reflection for small $n \geqslant 2$. If the dimension n of the search space is sufficiently large, then the magnitude $\widetilde{\beta}_i$ is also large so that $1/\widetilde{\beta}_i \to 0$. Therefore, the limiting case of the operator $\widetilde{R}_i^{-1}$ will be the operator of projection onto the hyperplane orthogonal to $\theta_i : I - \theta_i\theta_i^T$.

Above we analyzed the structure of the motion given by the Newtonian process (29). For an analysis of the whole picture it is necessary to study the structure of the vector field $\bar{y}(x)$ defined by formulas (34), (35), and (37).

The process of accumulating perspective directions (the organization of memory) for $\bar{y}(x)$ was considered in the beginning of the section. We can see from (37) that the motion generated by this vector field is also defined by the operator (39). In fact, for the Newtonian motion we have, in the case of K realizations θ_i, that

$$\widetilde{\bar{y}}(x) = \widetilde{\gamma}^{-1}(x)\prod_{i=1}^{K}\widetilde{R}_i^{-1}(x)\nabla\varphi(x), \tag{49}$$

where the elementary operator $\widetilde{R}_i(x)$ is defined in (47). In this case the motion (34) – (37) can also be represented in a form similar to (49)

$$\bar{y}(x) = \frac{1}{p(x)}\prod_{i=1}^{K}R_i(x)\nabla\varphi(x),$$

where

$$R_i(x) = \gamma(x)\left[I - \frac{cn}{2\gamma(x)p(x)}\nu_i\theta_i\theta_i^T\right]. \tag{50}$$

(see (37)).

Comparing (47) and (50), we see that these operators have a similar structure and therefore can be considered in the framework of the unified operator of statistical processing of information (39). At the same time we note that expression (34) can be written in the form

$$\bar{y}(x) = \gamma(x)\left[I + \frac{c}{2\gamma(x)p(x)}\nabla^2\varphi(x)\right]\frac{\nabla\varphi(x)}{p(x)}.$$

Hence, we can see that the motion (34)–(37) is a regularization of the Newtonian motion. The structure of the motion given by (34)–(37) is considered in more detail in [6, 7].

6. CONCLUSION

The above-considered structural properties of the motion generated by variation of the functional of the randomized problem make it possible to construct modules of nonlocal search. The usage of these modules leads to algorithms that can be interpreted as adaptive generalizations of the known deterministic methods (the method of variable polyhedron, the conjugate direction method, the ellipsoidal methods, etc.) [6, 7]. The algorithms suggested combine the learning process of the objective function and the process of its optimization. These algorithms, to the contrary of their traditional analogs, use only values of the objective function.

REFERENCES

1. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, New York (1981).
2. A. I. Kaplinskii and A. I. Propoi, "Methods of nonlocal search based on potential theory," *Avtomat. Telemekh.*, No. 7, 55–64 (1993).
3. A. I. Kaplinskii and A. I. Propoi, *The Variational Approach to Construction of Nonlocal Search Algorithms* [in Russian], VNIISI, Moscow (1986).
4. A. D. Myshkis, *Mathematics. Special Courses* [in Russian], Nauka, Moscow (1971).
5. V. S. Vladimirov, *Equations of Mathematical Physics* [in Russian], Nauka, Moscow (1988).
6. A. I. Kaplinskii and A. I. Propoi, *Structural Components of Methods for Nonlocal Search Using Potential Theory* [in Russian], VNIISI, Moscow (1990).
7. A. I. Kaplinskii and A. I. Propoi, *Development of Numerical Methods for Nonlocal Search Using Potential Theory* [in Russian], VNIISI, Moscow (1990).
8. I. M. Prudnikov, "The method of global optimization of a function and the estimate of its rate of convergence," *Avtomat. Telemekh.*, No.12, 72–81 (1993).
9. N. Z. Shor, *Methods for Minimization of Nondifferentiable Functions and Their Applications*, [in Russian], Naukova Dumka, Kiev (1979).
10. Ya. Z. Tsypkin, *Foundations of Information Theory of Identification* [in Russian], Nauka, Moscow, 1984.
11. B. T. Polyak and Ya. Z. Tsypkin, "Optimal pseudogradient algorithms," *Automat. Telemekh.*, No. 8, 74–84 (1980).
12. S. G. Mikhlin, *Linear Partial Differential Equations* [in Russian], Vysshaya Shkola, Moscow (1977).

APPENDIX

Proof of Proposition 1. We introduce the density function $p_\varepsilon(x) = \int_{R^n} p(x - \varepsilon y, y)\, dy$ of the random vector $X_\varepsilon = X + \varepsilon Y$. Considering this density as a function of ε

$$p_\varepsilon(x) \equiv h(x) = h(0) + h'(0)\varepsilon + \frac{1}{2}h''(0)\varepsilon^2 + o(\varepsilon^2),$$

we evaluate the corresponding derivatives of $h(\varepsilon)$ for $\varepsilon = 0$. We have for the first derivative

$$h'(\varepsilon) = \int_{R^n} \frac{d}{d\varepsilon} p(x - \varepsilon y, y)\, dy = -\int_{R^n} (\nabla_x p(x - \varepsilon y, y), y)\, dy = -\int_{R^n} \operatorname{div}_x [p(x - \varepsilon y, y) y]\, dy.$$

We use here the relation $\int_{R^n} (\nabla_x f(x), y)\, dy = \int_{R^n} \operatorname{div}_x [f(x) y]\, dy$, where $(\cdot, \cdot)$ is the inner product in R^n. (We also assume that the conditions for differentiability of the integral with respect to a parameter hold.) Hence,

$$h'(0) = -\int_{R^n} \operatorname{div}_x [p(x, y) y]\, dy = -\operatorname{div}_x [p(x) \int_{R^n} y p(y|x)\, dy] = -\operatorname{div}_x [p(x) \hat{y}(x)] \tag{A.1}$$

in accordance with (5) and (6). The same constructions are also used for the evaluation

$$h''(\varepsilon) = \frac{d^2}{d\varepsilon^2} \int_{R^n} p(x - \varepsilon y, y)\, dy.$$

By using (A.1) obtain that

$$\frac{d^2}{d\varepsilon^2} p(x - \varepsilon y, y) = \frac{d}{d\varepsilon} \left[\frac{d}{d\varepsilon} p(x - \varepsilon y, y) \right] = -\frac{d}{d\varepsilon} \operatorname{div}_x [p(x - \varepsilon y, y) y] =$$

$$= -\operatorname{div}_x \left[\frac{d}{d\varepsilon} p(x - \varepsilon y, y) \right] = -\operatorname{div}_x \{ -\operatorname{div}_x [p(x - \varepsilon y, y) y] \} = \operatorname{div}_x \operatorname{div}_x [p(x - \varepsilon y, y) y].$$

Therefore

$$h''(0) = \int_{R^n} \operatorname{div}_x \operatorname{div}_x [p(x, y) y]\, dy. \tag{A.2}$$

We transform (A.2) into a form similar to (A.1)

$$h''(0) = \operatorname{div}_x \int_{R^n} \operatorname{div}_x [p(x, y) y] y dy = \tag{A.3}$$

$$= \operatorname{div}_x \begin{pmatrix} \operatorname{div}_x \int_{R^n} p(x, y) y y_1\, dy \\ \cdots\cdots\cdots\cdots\cdots \\ \operatorname{div}_x \int_{R^n} p(x, y) y y_n\, dy \end{pmatrix} = \operatorname{div}_x \begin{pmatrix} \operatorname{div}_x [p(x) E[Y Y_1 | X = x]] \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \operatorname{div}_x [p(x) E[Y Y_n | X = x]] \end{pmatrix},$$

where

$$E[Y Y_i | X = x] = \int_{R^n} y y_i p(x, y)\, dy.$$

We now use the known relation

$$E[Y Y^T | X = x] = E[Y | X = x] E[Y | X = x]^T + K(x) = \hat{y}(x) \hat{y}^T(x) + K(x), \tag{A.4}$$

or, in coordinate form, $E[YY_i|X=x] = E[Y_i|X=x]E[Y_i|X=x] + K_i(x)$, where $K_i(x), i=1,\ldots,n$, is the i-th column of the covariance matrix $K(x)$ (7). By using (A.4) we rewrite (A.3) in the form

$$h''(0) = \operatorname{div}_x \begin{pmatrix} \operatorname{div}_x[p(x)\bar{y}(x)\bar{y}_1(x) + K_1(x)] \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \operatorname{div}_x[p(x)\bar{y}(x)\bar{y}_n(x) + K_n(x)] \end{pmatrix} = \tag{A.5}$$

$$= \operatorname{div}_x \begin{pmatrix} \operatorname{div}_x[p(x)\bar{y}(x)\bar{y}_1(x)] \\ \ldots\ldots\ldots\ldots\ldots \\ \operatorname{div}_x[p(x)\bar{y}(x)\bar{y}_n(x)] \end{pmatrix} + \operatorname{div}_x \begin{pmatrix} \operatorname{div}_x[p(x)K_1(x)] \\ \ldots\ldots\ldots\ldots \\ \operatorname{div}_x[p(x)K_n(x)] \end{pmatrix}.$$

The first term in (A.5) can be rewritten as

$$\operatorname{div}_x \begin{pmatrix} \operatorname{div}_x[p(x)\bar{y}(x)\bar{y}_1(x)] \\ \ldots\ldots\ldots\ldots\ldots \\ \operatorname{div}_x[p(x)\bar{y}(x)\bar{y}_n(x)] \end{pmatrix} = \operatorname{div}_x \left\{ \bar{y}(x)\operatorname{div}[p(x)\bar{y}(x)] + \nabla^T \bar{y}(x) p(x)\bar{y}(x) \right\},$$

where $\nabla\bar{y}(x)$ is a matrix with columns $\nabla\bar{y}_i(x), i=1,\ldots,n$. Denoting $h_i(x) = p(x)K_i(x)$ we obtain finally (18).

Proof of Proposition 4. The derivation of formula (29) is based on the following theorem from [12].

Theorem. If $\mu \in L_2(\Omega)$, then the generalized second-order derivatives of the function

$$\eta(x) = \int_\Omega \frac{\mu(\xi)\,d\xi}{r^{n-2}}, \quad r = \|x-\xi\|,$$

exist; they belong to $L_2(\Omega)$ and are written as

$$\frac{\partial^2 \eta(x)}{\partial x_i \partial x_j} = -\frac{(n-2)\omega_n}{n}\delta_{ij}\mu(x) + \int_\Omega \mu(\xi)\frac{\partial^2}{\partial x_i \partial x_j}\left(\frac{1}{r^{n-2}}\right) dxi, \tag{A.6}$$

where δ_{ij} is Kronecker's symbol.

By using (A.6) we obtain for (23) that

$$\frac{\partial^2 \varphi(x)}{\partial x_i \partial x_j} = \frac{\tilde{f}(x)p_U(x)}{n}\delta_{ij} + \int_{r>\rho} \tilde{f}(\xi)\frac{\partial^2 G(x,\xi)}{\partial \xi_i \partial \xi_j} p_U(\xi)\,d\xi +$$

$$\int_{r<\rho} \left[\tilde{f}(\xi)p_U(\xi) - \tilde{f}(x)p_U(x)\right]\frac{\partial^2 G(x,\xi)}{\partial \xi_i \partial \xi_j}\,d\xi,$$

where for $n > 2$

$$G(X,\xi) = -\frac{1}{(n-2)\omega_n \|x-\xi\|^{n-2}} = -\frac{1}{(n-2)\omega_n r^{n-2}},$$

$$\frac{\partial^2 G(x,\xi)}{\partial \xi_i \partial \xi_j} = \frac{\delta_{ij}}{\omega_n r^n} - \frac{n}{\omega_n r^{n+2}}(x_i - \xi_i)(x_j - \xi_j).$$

Hence,

$$\frac{\partial^2 \varphi(x)}{\partial x_i \partial x_j} = \tilde{f}(x)p_U(x)\frac{\delta_{ij}}{n} + \frac{\delta_{ij}}{\omega_n}\int_{r>\rho} \frac{\tilde{f}(\xi)p_U(\xi)}{r^n}\,d\xi - \frac{n}{\omega_n}\int_{r>\rho} \frac{\tilde{f}(\xi)}{r^n}\left(\frac{x_i-\xi_i}{r}\right)\left(\frac{x_j-\xi_j}{r}\right)p_U(\xi)\,d\xi +$$

$$+\frac{\delta_{ij}}{\omega_n}\int_{r<\rho} \frac{\tilde{f}(\xi)p_U(\xi) - \tilde{f}(x)p_U(x)}{r^n}\,d\xi - \frac{n}{\omega_n}\int_{r<\rho} \frac{\tilde{f}(\xi)p_U(\xi) - \tilde{f}(x)p_U(x)}{r^n}\left(\frac{x_i-\xi_i}{r}\right)\left(\frac{x_j-\xi_j}{r}\right)d\xi.$$

Using the notations (31) – (33) of Proposition 4, we finally obtain (30).